EQUIVALENT REPRESENTATIONS OF HIDDEN MARKOV MODELS

BERLIAN SETIAWATY

Department of Mathematics,
Faculty of Mathematics and Natural Sciences,
Bogor Agricultural University
Jl. Meranti, Kampus IPB Darmaga, Bogor, 16680 Indonesia

Abstract. In this article, we classify the class of hidden Markov models through the laws of the observation processes, since the Markov chains are not observable. Here, we also present some properties regarding this classification.
Key words: Hidden Markov, representations, equivalent.

1. Introduction

A hidden Markov model consists of two processes, ie: a Markov chain and an observed process. The Markov chain in a hidden Markov model is not observable. So, inference concerning the hidden Markov model has to be based on the information of the observed process alone. By this fact, two hidden Markov models wkich have the observed processes with the same law are theoretically *indistinguishable*. In this case, its is said that these two models are *equivalent* and their representations are also said to be *equivalent*.

In this article, we study the properties of this equivalence relation which give characteristics to the class of hidden Markov models.

In section 2, a hidden Markov is defined and an example is also given. In section 3, we define the equivalence relation in the class of hidden markov models and section 4 presents some properties of this equivalence relation.

2. Hidden Markov Models

Let $\{X_t : t \in \boldsymbol{N}\}$ be a finite state Markov chain defined on a probability space $(\Omega, \mathcal{F}, P)$. Suppose that $\{X_t\}$ is not observed directly, but rather there is an *observation* process $\{Y_t : t \in \boldsymbol{N}\}$ defined on $(\Omega, \mathcal{F}, P)$. Then consequently, the Markov chain is said to be *hidden* in the observations. A pair of stochastic processes $\{(X_t, Y_t) : t \in \boldsymbol{N}\}$ is called a hidden Markov model. Precisely, according to [3], a hidden Markov model is formally defined as follows.

Definition 2.1. A pair of discrete time stochastic processes $\{(X_t, Y_t) : t \in \boldsymbol{N}\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values in a set $\boldsymbol{S} \times \mathcal{Y}$, is said to be a ***hidden Markov model*** (HMM), if it satisfies the following conditions.

1. $\{X_t\}$ is a finite state Markov chain.
2. Given $\{X_t\}$, $\{Y_t\}$ is a sequence of conditionally independent random variables.
3. The conditional distribution of Y_n depends on $\{X_t\}$ only through X_n.
4. The conditional distribution of Y_t given X_t does not depend on t.

Assume that the Markov chain $\{X_t\}$ ***is not observable***. The cardinality K of $\boldsymbol{S}$, will be called the ***size*** of the hidden Markov model.

The following is an example of a hidden Markov model which is adapted from [1].

Example 2.2. Let $\{X_t\}$ be a Markov chain defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values on $\boldsymbol{S} = \{1, \dots, K\}$. The observed process $\{Y_t\}$ is defined by

$$Y_t = c(X_t) + \sigma(X_t)\omega_t, \qquad t \in \boldsymbol{N}, \tag{2.1}$$

where c and σ are real valued functions and positive real valued function on $\boldsymbol{S}$ respectively, and $\{\omega_t\}$ is a sequence of $N(0,1)$ independent, identically distributed (i.i.d.) random variables.

Since $\{\omega_t\}$ is a sequence of $N(0,1)$ i.i.d. random variables, then given $\{X_t\}$, $\{Y_t\}$ is a sequence of independent random variables. From (2.1), it is clear that Y_t is a function of X_t only, then the third condition of Definition 2.1 holds. Let $y \in \mathcal{Y}$ and $i \in \boldsymbol{S}$. Let $c_i = c(i)$ and $\sigma_i = \sigma(i)$, then

$$\begin{aligned} P(Y_t \le y | X_t = i) &= P(c_i + \sigma_i \omega_t \le y) \\ &= P(\sigma_i \omega_t \le y - c_i) \\ &= \int_{-\infty}^{y - c_i} \phi_i(z)\, dz, \end{aligned} \tag{2.2}$$

where

$$\phi_i(z) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z}{\sigma_i}\right)^2}. \tag{2.3}$$

Thus from (2.2) and (2.3), the conditional density of Y_t given $X_t = i$ is $\phi_i(\cdot - c_i)$ which does not depend on t. Therefore it can be concluded that $\{(X_t, Y_t)\}$ is a hidden Markov model.

3. Representations of a Hidden Markov Model

Let $\{(X_t, Y_t)\}$ be a hidden Markov model defined on the probability space $(\Omega, \mathcal{F}, P)$, taking values on $\boldsymbol{S} \times \mathcal{Y}$, where $\boldsymbol{S} = \{1, \dots, K\}$ and $\mathcal{Y} = \boldsymbol{R}$. Let $A = (\alpha_{ij})$ be the transition probability matrix and $\pi = (\pi_i)$ be the initial probability distribution of the Markov chain $\{X_t\}$. Assume for $i = i, \dots, K$, the conditional densities of Y_t given $X_t = i$ with respect to the measure μ, $p(\cdot|i)$, belong to the same family $\boldsymbol{\mathcal{F}}$, where $\boldsymbol{\mathcal{F}} = \{f(\cdot|\theta) : \theta \in \Theta\}$ is a family of densities on a Euclidean space with respect to the measure μ, indexed by $\theta \in \Theta$. This means that for each $i = 1, \dots, K$,

$$p(\cdot|i) = f(\cdot, \theta_i),$$

for some $\theta_i \in \Theta$.

From the result of [4], it can be seen that the law of the hidden Markov model $\{(X_t, Y_t)\}$ is completely specified by :
(a). The size K.
(b). The transition probability matrix $A = (\alpha_{ij})$, satisfying

$$\alpha_{ij} \geq 0, \qquad \sum_{j=1}^{K} \alpha_{ij} = 1, \qquad i, j = 1, \dots, K.$$

(c). The initial probability distribution $\pi = (\pi_i)$ satisfying

$$\pi_i \geq 0, \qquad i = 1, \dots, K, \qquad \sum_{i=1}^{K} \pi_i = 1.$$

(d). The vector $\theta = (\theta_i)^T$, $\theta_i \in \Theta$, $i = 1, \dots, K$, which desribes the conditional
densities of Y_t given $X_t = i$, $i = 1, \dots, K$.

Definition 3.1. Let $\phi = (K, A, \pi, \theta)$. The parameter ϕ is called a ***representation*** of the hidden Markov model $\{(X_t, Y_t)\}$.

Thus, the hidden Markov model $\{(X_t, Y_t)\}$ can be represented by a representation $\phi = (K, A, \pi, \theta)$.

On the otherhand, we can also generate a hidden Markov model $\{(X_t, Y_t)\}$ from a representation $\phi = (K, A, \pi, \theta)$, by choosing a Markov chain $\{X_t\}$ which takes values on $\{1, \dots, K\}$ and its law is determined by the $K \times K$-transition probability matrix A and the initial probability π, and an observation process $\{Y_t\}$ taking values on $\mathcal{Y}$, where the density functions of Y_t given $X_t = i$, $i = 1, \dots, K$ are determined by θ.

4. Equivalent Representations

Let $\phi = (K, A, \pi, \theta)$ and $\widehat{\phi} = (\widehat{K}, \widehat{A}, \widehat{\pi}, \widehat{\theta})$ be two representations which respectively generate hidden Markov models $\{(X_t, Y_t)\}$ and $\{(\widehat{X}_t, Y_t)\}$. The $\{(X_t, Y_t)\}$ takes values on $\{1, \dots, K\} \times \mathcal{Y}$ and $\{(\widehat{X}_t, Y_t)\}$ takes values on $\{1, \dots, \widehat{K}\} \times \mathcal{Y}$. For any $n \in \boldsymbol{N}$, let $p_\phi(\cdot, \cdots, \cdot)$ and $p_{\widehat{\phi}}(\cdot, \cdots, \cdot)$ be the n-dimensional joint density function of $Y_1, \dots Y_n$ with respect to ϕ and $\widehat{\phi}$. Suppose that for every $n \in \boldsymbol{N}$,

$$p_\phi(Y_1, \dots, Y_n) = p_{\widehat{\phi}}(Y_1, \dots, Y_n).$$

Then $\{Y_t\}$ has the same law under ϕ and $\widehat{\phi}$. Since in hidden Markov models $\{(X_t, Y_t)\}$ and $\{(\widehat{X}_t, Y_t)\}$, the Markov chains $\{X_t\}$ and $\{\widehat{X}_t\}$

are not observable and we only observed the values of $\{Y_t\}$, then theoretically, the hidden Markov models $\{(X_t, Y_t)\}$ and $\{(\widehat{X}_t, Y_t)\}$ are *indistinguishable*. In this case, it is said that $\{(X_t, Y_t)\}$ and $\{(\widehat{X}_t, Y_t)\}$ are *equivalent*. The representations ϕ and $\widehat{\phi}$ are also said to be *equivalent*, and will be denoted as $\phi \sim \widehat{\phi}$.

For each $K \in \boldsymbol{N}$, define

$$\begin{aligned}\Phi_K \;=\; \Big\{ \phi \,:\, \phi = (K, A, \pi, \theta), \text{ where } A, \pi \text{ and } \theta \text{ satisfy :} \\ A = (\alpha_{ij}), \quad \alpha_{ij} \geq 0, \quad \sum_{j=1}^{K} \alpha_{ij} = 1, \quad i, j = 1, \ldots, K \\ \pi = (\pi_i), \quad \pi_i \geq 0, \quad i = 1, \ldots, K, \quad \sum_{i=1}^{K} \pi_i = 1 \\ \theta = (\theta_i)^T, \quad \theta_i \in \Theta, \quad i = 1, \ldots, K \Big\}\end{aligned} \tag{4.1}$$

and

$$\Phi = \bigcup_{K \in \boldsymbol{N}} \phi_K. \tag{4.2}$$

The relation $\sim$ is now defined on Φ as follows.

Definition 4.1. Let $\phi, \widehat{\phi} \in \Phi$. Representations ϕ and $\widehat{\phi}$ are said to be ***equivalent***, denoted as

$$\phi \sim \widehat{\phi}$$

if and only if for every $n \in \boldsymbol{N}$,

$$p_\phi(Y_1, Y_2, \ldots, Y_n) = p_{\widehat{\phi}}(Y_1, Y_2, \ldots, Y_n).$$

Remarks 4.2. It is clear that relation $\sim$ forms an equivalence relation on Φ.

Let $\phi = (K, A, \pi, \theta) \in \Phi_K$, then under ϕ, $Y_1, \ldots, Y_n$, for any n, has joint density

$$p_\phi(y_1, \ldots, y_n) = \sum_{x_1=1}^{K} \cdots \sum_{x_n=1}^{K} \pi_{x_1} f(y_1, \theta_{x_1}) \cdot \prod_{t=2}^{n} \alpha_{x_{t-1}, x_t} f(y_t, \theta_{x_t}). \tag{4.3}$$

Let σ be any permutation of $\{1, 2, \ldots, K\}$. Define

$$\begin{aligned}\sigma(A) &= (\alpha_{\sigma(i), \sigma(j)}) \\ \sigma(\pi) &= (\pi_{\sigma(i)}) \\ \sigma(\theta) &= (\theta_{\sigma(i)})^T.\end{aligned}$$

Let

$$\sigma(\phi) = (K, \sigma(A), \sigma(\pi), \sigma(\theta)),$$

then $\sigma(\phi) \in \Phi_K$ and easy to see from (4.3) that

$$p_\phi(y_1, \ldots, y_n) = p_{\sigma(\phi)}(y_1, \ldots, y_n).$$

implying $\phi \sim \sigma(\phi)$. So we have the following lemma.

Lemma 4.3. *Let $\phi \in \Phi_K$, then for every permutation σ of $\{1, 2, \ldots, K\}$,*

$$\sigma(\phi) \sim \phi.$$

Lemma 4.4. *Let $\phi = (K, A, \pi, \theta) \in \Phi_K$. If $\theta_i = \gamma$, $i = 1, \ldots, K$, for some $\gamma \in \Theta$, then*

$$\phi \sim \phi(\gamma),$$

where $\phi(\gamma) = (1, (1), (1), (\gamma)) \in \Phi_1$.

Proof :
For any $n \in \boldsymbol{N}$,

$$\begin{aligned} p_\phi(y_1, \ldots, y_n) &= \sum_{x_1=1}^{K} \cdots \sum_{x_n=1}^{K} \pi_{x_1} f(y_1, \gamma) \prod_{t=1}^{n} \alpha_{x_{t-1}, x_t} f(y_t, \gamma) \\ &= \prod_{t=1}^{n} f(y_t, \gamma) \\ &= p_{\phi(\gamma)}(y_1, \ldots, y_n), \end{aligned}$$

where $\phi(\gamma) = (1, (1), (1), (\gamma)) \in \Phi_1$. Hence $\phi \sim \phi(\gamma)$. ■

Lemma 4.5. *Let $\phi = (K, A, \pi, \theta) \in \Phi_K$, where π is a stationary probability distribution of A. Let N be the number of non-zero π_i. Then there is $\widehat{\phi} = (N, \widehat{A}, \widehat{\pi}, \widehat{\theta}) \in \Phi_N$, such that :*

1. *$\widehat{\pi}_i > 0$, for $i = 1, \ldots, N$.*
2. *$\widehat{\pi}$ is a stationary probability distribution of $\widehat{A}$.*
3. *$\phi \sim \widehat{\phi}$.*

Proof :
Let $\phi = (K, A, \pi, \theta) \in \Phi_K$, where π is a stationary probability distribution of A. Let N be the number of non-zero π_i. Without lost of generality, suppose that

$$\pi_i = 0, \qquad \text{for } i = N+1, \ldots, K.$$

Since

$$\pi A = \pi,$$

then

$$\alpha_{ij} = 0, \qquad \text{for } i = 1, \ldots, N; \;\; j = N+1, \ldots, K.$$

Set

$$\begin{aligned} \widehat{\alpha}_{ij} &= \alpha_{ij}, \qquad i, j = 1, \ldots, N \\ \widehat{\pi}_i &= \pi_i, \qquad i = 1, \ldots, N \\ \widehat{\theta}_i &= \theta_i, \qquad i = 1, \ldots, N. \end{aligned}$$

Let

$$\widehat{A} = (\widehat{\alpha}_{ij}), \qquad \widehat{\pi} = (\pi_i), \qquad \widehat{\theta} = (\theta_i)^T$$

and $\widehat{K}$-row vector

$$\widehat{e} = (1, 1, \dots, 1).$$

Then it is clear that

$$\widehat{\pi}_i > 0, \qquad i = 1, \dots, N$$

and

$$\widehat{\pi}\,\widehat{A} = \widehat{\pi}.$$

Let

$$\widehat{B}(y) = \begin{pmatrix} f(y,\theta_1) & 0 & \cdots & 0 \\ 0 & f(y,\theta_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(y,\theta_N) \end{pmatrix}, \qquad y \in \mathcal{Y}$$

and

$$\widehat{M}(y) = \widehat{A}\widehat{B}(y).$$

Take $\widehat{\phi} = (N, \widehat{A}, \widehat{\pi}, \widehat{\theta})$, then $\widehat{\phi} \in \Phi_N$ and

$$\begin{aligned} p_\phi(y_1, \dots, y_n) &= \pi M(y_1) M(y_2) \cdots M(y_n) e \\ &= \widehat{\pi}\widehat{M}(y_1)\widehat{M}(y_2) \cdots \widehat{M}(y_n)\widehat{e} \\ &= p_{\widehat{\phi}}(y_1, \dots, y_n), \end{aligned}$$

implying $\phi \sim \widehat{\phi}$. ■

Next lemma gives sufficient conditions for representations to be equivalent. The idea of this lemma comes from [2]

Lemma 4.6. *Let $\phi = (K, A, \pi, \theta) \in \Phi_K$ and $\widehat{\phi} = (\widehat{K}, \widehat{A}, \widehat{\pi}, \widehat{\theta}) \in \Phi_{\widehat{K}}$. If there are $K \times \widehat{K}$-matrix X and $\widehat{K} \times K$-matrix Y, such that*

$$\begin{aligned} \widehat{A} &= YAX && (4.4) \\ X\widehat{B}(y) &= B(y)X, \qquad \forall y \in \mathcal{Y} && (4.5) \\ \widehat{\pi} &= \pi X \\ \widehat{e} &= Ye \\ XY &= I_K, \end{aligned}$$

then $\phi \sim \widehat{\phi}$.

Proof :
From (4.4) and (4.5), for every $y \in \mathcal{Y}$,

$$\begin{aligned}\widehat{M}(y) &= \widehat{A}\widehat{B}(y)\\ &= YAX\widehat{B}(y)\\ &= YAB(y)X\\ &= YM(y)X.\end{aligned}$$

For any $n \in \boldsymbol{N}$,

$$\begin{aligned}p_{\widehat{\phi}}(y_1,\dots,y_n) &= \widehat{\pi}\widehat{B}(y_1)\widehat{M}(y_2)\cdots\widehat{M}(y_n)\widehat{e}\\ &= \pi X\widehat{B}(y_1)YM(y_2)X\cdots YM(y_n)XYe\\ &= \pi B(y_1)XYM(y_2)X\cdots YM(y_n)XYe\\ &= \pi B(y_1)I_KM(y_2)\cdots I_KM(y_n)I_Ke\\ &= \pi B(y_1)M(y_2)\cdots M(y_n)e\\ &= p_\phi(y_1,\dots,y_n).\end{aligned}$$

Hence $\phi \sim \widehat{\phi}$. ■

Lemma 4.7. *Let $\phi = (K, A, \pi, \theta) \in \Phi_K$ and $\widehat{\phi} = (\widehat{K}, \widehat{A}, \widehat{\pi}, \widehat{\theta}) \in \Phi_{\widehat{K}}$, where π and $\widehat{\pi}$ are stationary probability distributions of A and $\widehat{A}$ respectively. If there are $K \times \widehat{K}$-matrix X and $\widehat{K} \times K$-matrix Y, such that*

$$\begin{aligned}\widehat{M}(y) &= YM(y)X, \qquad \forall y \in \mathcal{Y} \qquad (4.6)\\ \widehat{\pi} &= \pi X\\ \widehat{e} &= Ye\\ XY &= I_K,\end{aligned}$$

then $\phi \sim \widehat{\phi}$.

Remarks 4.8. Equation (4.6) implies $\widehat{A} = YAX$.

Proof :
For any $n \in \boldsymbol{N}$,

$$\begin{aligned}p_{\widehat{\phi}}(y_1,\dots,y_n) &= \widehat{\pi}\widehat{M}(y_1)\widehat{M}(y_2)\cdots\widehat{M}(y_n)\widehat{e}\\ &= \pi XYM(y_1)XYM(y_2)X\cdots YM(y_n)XYe\\ &= \pi I_KM(y_1)I_KM(y_2)I_K\cdots I_KM(y_n)I_Ke\\ &= \pi M(y_1)M(y_2)\cdots M(y_n)e\\ &= p_\phi(y_1,\dots,y_n).\end{aligned}$$

Thus $\phi \sim \widehat{\phi}$. ■

Based on Lemma 4.6 and Lemma 4.7, we derive the following lemmas.

Lemma 4.9. *For any $K \in \boldsymbol{N}$ and $\phi \in \Phi_K$, there is $\widehat{\phi} \in \Phi_{K+1}$, such that $\phi \sim \widehat{\phi}$.*

Proof :
Let $\phi = (K, A, \pi, \theta) \in \Phi_K$. Define a $K \times (K+1)$-matrix X and a $(K+1) \times K$-matrix Y respectively as follow

$$X = \begin{pmatrix} I_{K-1} & O_{K-1,2} \\ O_{1,K-1} & a \quad b \end{pmatrix}, \qquad Y = \begin{pmatrix} I_{K-1} & O_{K-1,1} \\ O_{2,K-1} & \begin{matrix} 1 \\ 1 \end{matrix} \end{pmatrix} \tag{4.7}$$

where a and b are any real numbers, such that $a, b \geq 0$ and $a + b = 1$. Notice that

$$XY = I_K$$

and

$$\widehat{e} = Ye.$$

Let $\widehat{A} = (\widehat{\alpha}_{ij})$ be a $(K+1) \times (K+1)$-matrix defined by

$$\begin{aligned} \widehat{A} &= YAX \\ &= \begin{pmatrix} \alpha_{1,1} & \cdots & \alpha_{1,K-1} & a \cdot \alpha_{1,K} & b \cdot \alpha_{1,K} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \alpha_{K-1,1} & \cdots & \alpha_{K-1,K-1} & a \cdot \alpha_{K-1,K} & b \cdot \alpha_{K-1,K} \\ \alpha_{K,1} & \cdots & \alpha_{K,K-1} & a \cdot \alpha_{K,K} & b \cdot \alpha_{K,K} \\ \alpha_{K,1} & \cdots & \alpha_{K,K-1} & a \cdot \alpha_{K,K} & b \cdot \alpha_{K,K} \end{pmatrix} \end{aligned} \tag{4.8}$$

It is clear that

$$\widehat{\alpha}_{ij} \geq 0, \qquad i, j = 1, \ldots, K+1$$

$$\sum_{j=1}^{K+1} \widehat{\alpha}_{ij} = 1, \qquad i = 1, \ldots, K+1.$$

Let $\widehat{\pi} = (\widehat{\pi}_i)$ be a $(K+1)$-row vector which is defined by

$$\begin{aligned} \widehat{\pi} &= \pi X \\ &= (\pi_1, \ldots, \pi_{K-1}, a \cdot \pi_K, b \cdot \pi_K), \end{aligned} \tag{4.9}$$

then it is obvious that

$$\widehat{\pi}_i \geq 0, \qquad i = 1, \ldots, K+1$$

$$\sum_{i=1}^{K} \widehat{\pi}_i = 1.$$

Let $\widehat{\theta} = (\widehat{\theta}_i)$ be a $K+1$-column vector which is defined by

$$\begin{aligned} \widehat{\theta} &= Y\theta \\ &= (\theta_1, \ldots, \theta_{K-1}, \theta_K, \theta_K)^T \end{aligned} \tag{4.10}$$

and for $y \in \mathcal{Y}$, $\widehat{B}(y)$ be a $(K+1)\times(K+1)$-diagonal matrix defined by

$$\widehat{B}(y) = \begin{pmatrix} f(y,\theta_1) & 0 & \cdots & 0 & 0 \\ 0 & f(y,\theta_2) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & f(y,\theta_K) & 0 \\ 0 & 0 & \cdots & 0 & f(y,\theta_K) \end{pmatrix}. \qquad (4.11)$$

Notice that

$$X\widehat{B}(y) = B(y)X, \qquad \forall y \in \mathcal{Y}.$$

Let $\widehat{\phi} = (K+1, \widehat{A}, \widehat{\pi}, \widehat{\theta})$, then $\widehat{\phi} \in \Phi_{K+1}$ and by Lemma 4.6, $\phi \sim \widehat{\phi}$. ■

From the proof of Lemma 4.9, for $\phi \in \Phi_K$, there are infinitely many $\widehat{\phi} \in \Phi_{K+1}$ depending on how a and b were chosen, such that $\phi \sim \widehat{\phi}$. So we have the following corollary.

Corollary 1. For $\phi \in \Phi_K$, there are infinitely many $\widehat{\phi} \in \Phi_{K=1}$ such that $\phi \in \widehat{\phi}$.

Lemma 4.10. *For any $K \in \boldsymbol{N}$ and $\phi = (K, A, \pi, \theta) \in \Phi_K$, where π is a stationary probability distribution of A, then there is $\phi = (K+1, \widehat{A}, \widehat{\pi}, \widehat{\theta}) \in \Phi_{K+1}$ such that :*

1. *$\widehat{\pi}$ is a stationary distribution of $\widehat{A}$.*
2. *$\widehat{\phi} \sim \phi$.*

Proof :
Let $\phi = (K, A, \pi, \theta) \in \Phi_k$, where π is a stationary probability distribution of A. Let $\widehat{\phi} = (K+1, \widehat{A}, \widehat{\pi}, \widehat{\theta}) \in \Phi_{K+1}$ as in the proof of Lemma 4.9, hence $\phi \sim \widehat{\phi}$. Since π is a stationary probability distribution of A, then

$$\pi A = A,$$

implying

$$\begin{aligned} \widehat{\pi}\widehat{A} &= \pi XYAX \\ &= \pi I_K AX \\ &= \pi AX \\ &= \pi X \\ &= \widehat{\pi}. \end{aligned}$$

So $\widehat{\pi}$ is a stationary probability distribution of $\widehat{A}$. ■

Remarks 4.11. In Lemma 4.10, if $\pi_i > 0$, for $i = 1, \ldots, K$, then by choosing $a, b > 0$ in matrix X, we have $\widehat{\pi}_i > 0$, for $i = 1, \ldots, K+1$.

Let $\phi = (K, A, \pi, \theta) \in \Phi_K$, then by Lemma 3.4 of [4], the conditional density function of $Y_1, \dots, Y_n$, given $X_1 = i$, $i = 1, \dots, K$, under ϕ is,

$$p_\phi(Y_1, \dots, Y_n | X_1 = i) = e_i^T B(Y_1) M(Y_2) \cdots M(Y_n) e.$$

Define,

$$q_\phi(Y_1, \dots, Y_n) = \max_{1 \le K} p_\phi(Y_1, \dots, Y_n | X_1 = i).$$

Lemma 4.12. *For any $K \in \boldsymbol{N}$ and $\phi \in \Phi_K$, there is $\widehat{\phi} \in \Phi_{K+1}$, such that :*

1. $\phi \sim \widehat{\phi}$.
2. $q_\phi(Y_1, \dots, Y_n) = q_{\widehat{\phi}}(Y_1, \dots, Y_n), \qquad$ *for every* $\quad n \in \boldsymbol{N}$.

Proof :
Let $\phi = (K, A, \pi, \theta) \in \Phi_K$. Let $\widehat{\phi} = (K+1, \widehat{A}, \widehat{\pi}, \widehat{\theta}) \in \Phi_{K+1}$ as in the proof of Lemma 4.9, then $\widehat{\phi} \sim \phi$. Notice that from definition of X in (4.7),

$$\widehat{e}_i^T = e_i^T X, \qquad \text{for} \quad \text{i= 1, \dots , K-1.} \tag{4.12}$$

Therefore by (4.12) and Lemma 3.4 of [4], for $i = 1, \dots, K-1$,

$$\begin{aligned} p_{\widehat{\phi}}(Y_1, \dots, Y_n | X_1 = i) &= \widehat{e}_i^T \widehat{B}(Y_1) \widehat{M}(Y_2) \cdots \widehat{M}(Y_n) \widehat{e} \\ &= e_i^T X \widehat{B}(Y_1) Y M(Y_2) X \cdots Y M(Y_n) X Y e \\ &= e_i^T B(Y_1) X Y M(Y_2) X \cdots Y M(Y_n) X Y e \\ &= e_i^T B(Y_1) I_K M(Y_2) I_K \cdots I_K M(Y_n) I_K e \\ &= e_i^T B(Y_1) M(Y_2) \cdots M(Y_n) e \\ &= p_\phi(Y_1, \dots, Y_n | X_1 = i) \end{aligned} \tag{4.13}$$

Since by (4.8), the K-th and $K+1$-th rows of $\widehat{A}$ are the same and $\widehat{\theta}_K = \widehat{\theta}_{K+1}$, then by Lemma 3.4 of [4],

$$\begin{aligned} & p_{\widehat{\phi}}(Y_1, \dots, Y_n | X_1 = K) \\ &= f(Y_1, \widehat{\theta}_K) \sum_{x_2=1}^{K+1} \cdots \sum_{x_n=1}^{K+1} \widehat{\alpha}_{K, x_2} f(Y_2, \widehat{\theta}_{x_2}) \prod_{t=3}^{n} \widehat{\alpha}_{x_{t-1}, x_t} f(Y_t, \widehat{\theta}_{x_t}) \\ &= f(Y_1, \widehat{\theta}_{K+1}) \sum_{x_2=1}^{K+1} \cdots \sum_{x_n=1}^{K+1} \widehat{\alpha}_{K+1, x_2} f(Y_2, \widehat{\theta}_{x_2}) \prod_{t=3}^{n} \widehat{\alpha}_{x_{t-1}, x_t} f(Y_t, \widehat{\theta}_{x_t}) \\ &= p_{\widehat{\phi}}(Y_1, \dots, Y_n | X_1 = K+1). \end{aligned} \tag{4.14}$$

Also notice that for $a, b \ge 0$ and $a + b = 1$,

$$a \widehat{e}_K^T + b \widehat{e}_{K=1}^T = e_K^T X, \tag{4.15}$$

then by (4.15) and Lemma 3.4 of [4],

$$\begin{aligned}
& ap_{\widehat{\phi}}(Y_1,\dots,Y_n|X_1=K)+bp_{\widehat{\phi}}(Y_1,\dots,Y_n|X_1=K+1)\\
&= a\widehat{e}_K^T\widehat{B}(Y_1)\widehat{M}(Y_2)\cdots\widehat{M}(Y_n)\widehat{e}+b\widehat{e}_K^T\widehat{B}(Y_1)\widehat{M}(Y_2)\cdots\widehat{M}(Y_n)\widehat{e}\\
&= \left(a\widehat{e}_K^T+b\widehat{e}_K^T\right)\widehat{B}(Y_1)\widehat{M}(Y_2)\cdots\widehat{M}(Y_n)\widehat{e}\\
&= e_K^T X\widehat{B}(Y_1)YM(Y_2)X\cdots YM(Y_n)XYe\\
&= e_K^T B(Y_1)XYM(Y_2)X\cdots YM(Y_n)XYe\\
&= e_K^T B(Y_1)I_K M(Y_2)I_K\cdots I_K M(Y_n)I_K e\\
&= e_K^T B(Y_1)M(Y_2)\cdots M(Y_n)e\\
&= p_\phi(Y_1,\dots,Y_n|X_1=K). \qquad (4.16)
\end{aligned}$$

Since $a, b \geq 0$ and $a+b=1$, then from (4.14) and (4.16),

$$p_{\widehat{\phi}}(Y_1,\dots,Y_n|\hat{X}=i)=p_\phi(Y_1,\dots,Y_n|X_1=K), \qquad \text{for } i=K,K+1. \qquad (4.17)$$

From (4.13) and (4.17),

$$\begin{aligned}
q_\phi(Y_1,\dots,Y_n) &= \max_{1\leq i\leq K} p_\phi(Y_1,\dots,Y_n|X_1=i)\\
&= \max_{1\leq i\leq K+1} p_{\widehat{\phi}}(Y_1,\dots,Y_n|X_1=i)\\
&= q_{\widehat{\phi}}(Y_1,\dots,Y_n)
\end{aligned}$$

■

By Lemma 4.9, we can define an order $\prec$ in $\{\Phi_K\}$.

Definition 4.13. Define an ***order*** $\prec$ on $\{\Phi_K\}$ by

$$\Phi_K \prec \Phi_L, \qquad K, L \in \boldsymbol{N},$$

if and only if for every $\phi \in \Phi_K$, there is $\widehat{\phi} \in \Phi_L$ such that $\phi \sim \widehat{\phi}$.

As a consequence of Lemma 4.9, Lemma 4.14 follows.

Lemma 4.14. *For every $K \in \boldsymbol{N}$,*

$$\Phi_K \prec \Phi_{K+1}.$$

From Lemma 4.14, the families of hidden Markov models represented by $\{\Phi_K\}$ are ***nested families*** as shown in Figure 1.

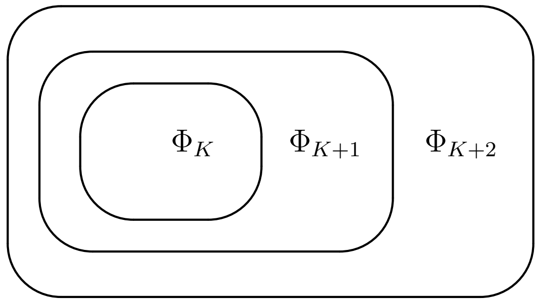

Figure 1.

References

[1] R.J. Elliott, L. Aggoun, and J.B. Moore. 1993. *Hidden Markov models: estimation and control*, Volume 29 of Application of Mathematics. Springer Verlag, New York.

[2] L, Finesso. 1990. *Consistent estimation of the order for hidden Markov models*. Ph D Thesis, Graduate School of the University of Maryland.

[3] T, Ryden. 1996. On recursive estimation for hidden Markov models. *Stochastic Processes and their Applications*, 66, 79-96.

[4] B. Setiawaty. 2002. A Hidden Markov Model: dependencies between random variables and its representations. *Jurnal Matematika dan Aplikasinya*, Vol. 1, No. 2.